



SUSMAN GODFREY L.L.P.

A REGISTERED LIMITED LIABILITY PARTNERSHIP



July 22, 2024

VIA ECF

Hon. Sidney H. Stein
 Daniel Patrick Moynihan
 United States Courthouse
 500 Pearl St.
 New York, NY 10007-1312

RE: *Authors Guild et al. v. OpenAI, Inc., et al., and Alter et al. v. OpenAI Inc., et al.*, Nos. 1:23-cv-08292-SHS & 1:23-cv-10211-SHS

Dear Judge Stein:

Pursuant to Rule 2(G) of Your Honor's Individual Practices, Plaintiffs request an informal discovery conference concerning Microsoft's refusal to search for documents responsive to twelve of Plaintiffs' Requests for Production of Documents ("RFPs"). These RFPs all seek material related to the scope of Microsoft's notice of and knowledge in connection with OpenAI's training of the Large Language Models ("LLMs") at issue in the case. These questions are at the center of this litigation and Microsoft should be compelled to search for materials that bear directly on them.

Background. Plaintiffs' First Amended Complaint alleges that OpenAI has directly violated Plaintiffs' copyrights in certain works of fiction and non-fiction ("Class Works") in the course of training its LLMs. The Complaint further alleges that Microsoft knew of, participated in, and enabled these violations as part of the companies' "joint project" to develop, integrate, and deploy these LLMs. *See* Dkt. 69 at ¶ 138-41. Specifically, Count I alleges that both OpenAI and Microsoft have committed direct copyright infringement of the Class Works, Count II alleges that two OpenAI entities have committed vicarious infringement, and Count III alleges that Microsoft and the OpenAI entities have committed contributory infringement. *See id.* at pp. 63-66.

On January 24, 2024, Plaintiffs served Microsoft with their first set of RFPs. Ex. 1. Many of the RFPs seek material that relates—in a variety of ways—to the training of OpenAI's LLMs. For example, RFP Nos. 17-18 request documents relating to "the use of commercial works of [fiction/non-fiction] to train" OpenAI's LLMs. *Id.* Such material is directly relevant to core factual questions of this litigation—e.g., whether OpenAI used commercial works to train its LLMs, what Microsoft knew about the training, when Microsoft knew it, and how Microsoft enabled this training.

Hon. Sidney H. Stein
 July 22, 2024
 Page 2

On February 23, 2024, Microsoft served responses asserting the general objection that it “will not search for or produce documents” if it interprets the request to “imply, suggest and/or assume that Microsoft trained or otherwise developed OpenAI’s models.” Ex. 2 at 5-6. Reiterating this objection, Microsoft indicated that it would not search for documents responsive to twelve RFPs because—based on an investigation it carried out in a different litigation—“Microsoft did not train ChatGPT.” Ex. 2 at RFP Nos. 17-18; 21-25; 27-29; 39-40.

Plaintiffs promptly raised the issue of this improper objection/response and the parties met and conferred by phone. The parties exchanged additional correspondence (Exs. 3-4) culminating with Plaintiffs’ June 13, 2024 letter explaining that Microsoft’s objection is “improper and does not supply a valid basis (e.g. relevance, burden, proportionality) for refusing Plaintiffs’ requests” and that “to the extent Microsoft maintains its refusal to search for relevant material responsive to these requests—because *inter alia* the requests ‘imply’ Microsoft developed OpenAI’s LLMs—the parties are at an impasse.” Ex. 5 at 2. Plaintiffs inquired of Microsoft whether further conferral was warranted and received no response.

Discussion. Microsoft refuses to search for material responsive to Plaintiffs’ RFPs on the basis that “since Microsoft did not train ChatGPT, it has no documents responsive to these requests.” *See* Ex. 4 at 2. This objection is illogical. As in other cases involving contributory as well as direct liability, Microsoft can possess documents *e.g.*, “relating to the use of commercial works of fiction” to train OpenAI’s LLMs (Ex. 1) without having done the training itself.¹

Moreover, Microsoft’s claim that it did not *train* these LLMs does not dispose of Plaintiffs’ allegation of direct infringement in count I of the First Amended Complaint. If, in the course of its work on the companies’ “joint project,” Dkt. 69 at ¶ 140, to develop these LLMs, Microsoft made copies of training data containing unauthorized copies of Class Works to, *e.g.*, placed on its servers, analyzed, or optimized, such actions would be relevant to Plaintiffs’ allegation of direct infringement. That would be true even if, on Microsoft’s apparently cramped understanding of the term, Microsoft did not undertake the specific actions it characterizes as “training.”

Critically, Microsoft’s objection—that, factually, it did not train OpenAI’s models—is also an invalid basis for refusing Plaintiffs’ discovery requests. “If a party objects to discovery requests, that party bears the burden of showing how... despite the broad and liberal construction afforded the federal discovery rules, each request is not relevant, or how each request is overly broad, burdensome, or oppressive.” *Financial Guaranty Insurance Company v. Putnam Advisory*

¹ As noted above, Microsoft makes the same argument in an alternative guise by asserting it searched for documents but did not locate any. However, what Microsoft claims is that it searched only for material addressing *whether Microsoft did the training itself*, not Plaintiffs requested information, including the extent of Microsoft’s knowledge of and contributions to the training. Thus, whether Microsoft defends its position by claiming it refuses to search, or whether it defends its position by saying it already conducted searches (but not for what Plaintiffs requested), the outcome is the same.

Hon. Sidney H. Stein
 July 22, 2024
 Page 3

Company, LLC, 314 F.R.D. 85, 87–88 (S.D.N.Y. 2016) (citations and marks omitted). Microsoft’s reliance on an objection—which is not grounded in relevance, burden or oppression—dooms their ability to carry this burden.

There is no dispute that the material sought bears on core questions of the litigation and is highly relevant to Plaintiffs’ claims. Microsoft has not identified *any* undue burden, let alone a burden that would render production of this highly relevant material disproportionate to the needs of the case. *See In re Zimmer M/L Taper Hip Prosthesis or M/L Taper Hip Prosthesis with Kinectiv Tech. & Versys Femoral Head Prod. Liab. Litig.*, No. 18-MC-2859 (PAC), 2020 WL 1812801, at *2 (S.D.N.Y. Apr. 9, 2020) (“Proportionality, which focuses on the marginal utility of the discovery sought, goes hand-in-hand with relevance, such that the greater the relevance of the information in issue, the less likely its discovery will be found to be disproportionate”).

Finally, Microsoft already produced material that supports Plaintiffs’ allegations. For example, its [REDACTED] (MSFT

AICPY_000004210 at 13) and that “[REDACTED]

” *Id.* at 12. Plaintiffs’ belief that additional discovery is necessary is based on facts as well as allegations.

Plaintiffs respectfully request an order directing Microsoft to search for and produce all non-privileged material responsive to RFP Nos. 17-18; 21-25; 27-29; and 39-40.

Respectfully,

LIEFF CABRASER HEIMANN SUSMAN GODFREY LLP COWAN, DEBAETS,
 & BERNSTEINS LLP ABRAHAMS & SHEPPARD
 LLP

/s/ Rachel Geman
 Rachel Geman

/s/ Rohit Nath
 Rohit Nath

/s/ Scott J. Sholder
 Scott J. Sholder

cc: All Counsel of Record (via ECF)